CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Design

The research used is Ex-post facto research. Ex post facto research is a model of research on the variables that happened before the research is conducted (Arikunto 2010: 17). Based on the level of explanation (level of explanation of the position of variables) this research is associative causal, that is a research is used to find the influence of independent variable to dependent variable (Sugiyono, 2010: 11). The study was conducted to determine the effect of gross domestic product and provincial minimum wage on labor absorption. This research uses quantitative approach and panel data analysis. The data used are six provincial panel data in Indonesia from 2011-2015. Panel observation and retrieval is useful in analyzing the dynamics of changes in labor absorption and factors which are closely related to employment in thirty-three Provinces in Indonesia over time.

3.2 Definition of Operasional Variable

3.2.1 Dependent Variable

a. Employment (Y)

Employment absorption is the number of employment that has been filled which is reflected from the number of labor force. The unit used is the soul. Labor absorption data is obtained from

Central Buureau of Statistics Publication of Labor Force Situation of six province in Java 2011-2015.

3.2.2 Independent Variable

a. Gross Regional Domestic Product (X1)

Gross regional domestic product is the total added value of goods and services produced by various sectors in each province in Indonesia within one year. GRDP in this research uses data of GRDP based on constant price 2010. Unit that is used is trillions rupiah. GRDP data is based on current price which is obtained from Central Bureau of Statistics on 2011 by online publication.

b. Province Minimum Wages (X2)

The provincial minimum wage is the lowest monthly wage consisting of basic wages including fixed allowances set by the provincial government in Indonesia. The unit used is rupiah. Provincial minimum wage data is obtained from Central Bureau of Statistics on 2011 by online publication.

c. Education (X3)

Educational data uses the number of workers completing their schooling by level, from elementary, junior high, high school, diploma or university. Education data is obtained from Central Bureau of Statistics on 2011 by online publication.

3.3 Collecting Data Technique

Data collection techniques used in this study is documentation study. Documentation technique is a technique of collecting data from various sources that are written. This study used documents issued by the Indonesian Central Bureau of Statistics such as six Provinces in figures and labor force catalog. The document will provide the data on gross regional domestic product, minimum wages rate and education of the six Provinces in Java in the period of 2011 until 2015.

3.4 Data type and Data Source

The data used in this study is secondary data taken from the publication of the Central Bureau of Statistics (CBS). The types of data and data sources used in this study are described in more detail in the following research table.

Table 4.
Data Description

	Data	**Sources**
Independent Variable	GRDP at constant prices	Online publication of Central Bureau of Statistics (CBS). 2011-2015
	Province Minimum Wages	Online publication of Central Bureau of Statistics (CBS). 2011 - 2015
	Education	Online publication of Central Bureau of Statistics (CBS). 2011 - 2015
Dedependent variable	Employment	Central Bureau of Statistics (CBS) catalogue, the workforce condition of each provinces in Indonesia 2011-2015

3.5 Method of Analysis

The data analysis used in this research is panel data estimation. Panel data is an econometric model that unifies between time series and cross section data, so that in panel data the number of observations is the result of observation time series (T> 1) with cross section (N> 1) observation. In performing the analysis, panel data can be divided into two namely the balance panel data and unbalance panel data. The balance of data panels occurs when the length of time for each cross section unit is the same. While unbalanced panel data occurs if the length of time is not the same for each unit cross section (Gujarati, 2012: 238). Through repeated observation of the cross section, panel data analysis allows one to study the dynamics of change with time series data. Therefore, panel data can explain two kinds of information that is cross section information on the differences between subjects and time series information reflecting the changes in time subject.

The combination of time series and cross section data can improve the quality and quantity of data with an approach that is not possible by using just one of them. Panel data analysis can study a group of subjects if the researcher wants to consider both dimensions of the data and the dimensions of time. According to Baltagi (in Gujarati, 2012: 237) the advantages of panel data are as follow:

1. Panel data estimation techniques can overcome heterogeneity.
2. By combining time series and cross section observations, panel data provides more information, more variety, less collinearity between variables and more efficiently.

3. Panel data is best suited for studying the dynamics of change.
4. Panel data is best for detecting and measuring impacts that simply can not be seen from pure cross-section or pure time series.
5. Panel data makes it easy to learn complex behavioral models.
6. Panel data can minimize the bias that can occur if the individual aggregates into a large aggregation.

The general equation for panel data estimation is as follows:

$$\mathbf{Y_{it} = \beta_0 + \beta_1 X_{1it} + e_{it}, \quad i = 1, 2, 3, ..., N ; t = 1, 2, ..., T}$$

where:

N : sum of the observation

T : time

N x T : panel data

To know the effect of free variable (GRDP, minimum wages rate and education) to dependent variable (employment) of six provinces in Java, researchers used multiple regression analysis methods of data panels. The formulation of this research model refers to the research of Dimas and Woyanti (2009) which analyzes the factors of employment in Jakarta and theoretical review. Here is the equation estimation model in this study:

$$\mathbf{EMP_{it} = \beta_0 + \beta_1 GRDP_{it} + \beta_2 PMW_{it} + \beta_3 EDU_{it} + e_{it}}$$

Details:

EMP = The absorption of labor force (people)

t = Research year 2011-2015

i = Province

β_0 = intersept (constanta)

$\beta_1, \beta_2, \beta_3, \beta_4$ = regression coefficient of each variables

PDRB = Gross National product (triliun rupiah)

PMW = provincial minimum wages (rupiah)

EDU = Education (people)

e = *error*

To achieve the research objectives, the data analysis in this research will be done through econometric model with the help of Eviews 7 program, while the stages of analysis are as follows:

3.5.1 Descriptive Statistics

Descriptive statistics is statistics that serves to describe or give an idea of the object under study through sample data or population without doing analysis and making conclusions that apply to the public (Sugiyono, 2012: 29).

3.5.2 The Choice of Estimation Method of panel Data

a. Panel Data Estimation Method

1) *Pooled Least Square Method*

The simplest approach in panel data processing is the ordinary least squares method applied in pool-shaped data. This method ignores the existence of individual and time dimensional differences (intersept and slope are considered equal / constant.

2) *Fixed Effect*

The fixed effect model is a model that assumes a constant slope coefficient but intercept varies between individuals.

3) *Random Effect*

In a random effect model, different parameters between regions and between times are included in the error. Every individual has a diversity of constants and applies to the observations within the individual.

b. Choice of Mode

1) *Chow Test*

Testing is conducted to choose whether the model will be analyzed using the common effect or fixed effect. It can be doneusing Chow Test. The hypothesis used is as follows:

Ho: Common Effect

Ha: Fixed Effect

The basis of rejection of the null hypothesis (ho) is when the probability value $F < \alpha$ (0.05).

2) *Hausman Test*

Testing to choose whether the model will be analyzed using random effect or fixed effect can be done using Hausman test. The hypothesis used in the Hausman test is as follows:

Ho: Random Effect

Ha: Fixed Effect

The basis of rejection of the null hypothesis (Ho) is when the probability value $F < \alpha$ (0.05).

3.6 Classical Assumption Test

3.6.1 Normality Test

Testing the normality of data is the test about the normal distribution of data. Normality testing is done with the intention to see whether or not the data is normal or close to normal. A good regression model has normal or near-normal data distribution. Normal distributed residual values can be known from the shape of the curves that make up the bell picture whose sides are widened until they are not finished. In addition to using the graph, the normality test can also be done by the Jarque-Bera method (JB test). The JB test is performed by looking at the Jarque-Bera probability value. According to Winarno (2015: 5.41) the normal distributed regression model has a probability JB > 0.05 ($\alpha = 0.05$). On the other hand, if the probability value < 0.05 then the data is not normally distributed.

3.6.2 Multicolinearity

Multicollinearity test is a test used to see the correlation between each independent variable. One method that can be used to determine the presence or absence of multicollinearity can be seen from the correlation value between two independent variables. If the correlation value is less than 0.8 then the independent variable does not have multicollinearity problem, and vice versa.

3.6.3 Heteroscedasticity

Heteroscedasticity is the situation of uneven or unequal data distribution of variance so that the test of significance is not valid. The heteroscedasticity test aims to find out whether in a regression model there is a residual variation inequality from one observation to another. If the residual variant from one observation to another observes remains, it is called heterocedasticity (same variant). One way to detect heteroscedasticity problems is to use the Glejser test. The Glejser test is performed by regressing all independent variables to the residual absolute value (Winarno, 2015: 5.16). if the probability value of free variable <0.05 (significant level or $\alpha = 0.05$) then heteroscedasticity occurs. On the other hand, if probability value> 0.05 then homocedasticity occurs.

3.6.4 Autocorrelation

Autocorrelation is the correlation between a series of observation members sorted by time series. According to Gujarati (2006: 37), the most popular test for detecting autocorrelation is the Durbin-Watson statistical test. Decision-making on this assumption requires two auxiliary values obtained from the Durbin-Watson table, dL and dU, with K = the number of independent variables and n = sample size. Testing is done by looking at the value of Durbin-Watson.

Table 5.
Rules of Auto Correlations

Null Hypothesis	Decision	If
There is no positive autocorrelation	Rejected	$0 < d < dL$
There is no positive autocorrelation	No decision	$dL \leq d \leq dU$
There is no negative autocorrelation	Rejected	$4 - dL < d < 4$
There is no negative autocorrelation	No decision	$4 - dU \leq d \leq 4 - dL$
There is no both positive and negative autocorrelation	Accepted	$dU < d < 4 - dU$

Source: Sofyan Yamin, 2011

3.7 Data Analysis Using Statistics Method

In testing hypothesis, some tests namely other individually regression coefficient test (t-test), complete regression coefficient test (F test), coefficient of determination test (R^2).

3.7.1 Partial Regression Test (t test)

Regression coefficient is used to determine the influence of independent variables partially to dependent variable. T test can be done by comparing probability value with significance level. If the value of Prob $<\alpha$ then the coefficient of the variable significantly affects the dependent variable and vice versa. Testing of regression result is done by using t test at 95% confidence degree or $\alpha = 5\%$ with the following conditions:

If the probability value t-statistics > 0.05 then H0 is rejected

If the probability t-statistic value < 0.05 then Ha is rejected

3.7.2 Simultaneous Test (F test)

F test (Simultaneous Test) is used to show whether the whole varable is independent of the dependent variable. The F test is also called the model feasibility test used to identify if the estimated regression model is feasible or not. Worthy here means that the estimated model is worthy of use to explain the effect of independent variables on the dependent variable. Test F can be done by looking at Prob (F-statistic). If the value of Prob (F-statistic) <0005 ($\alpha = 0.05$) then the regression coefficient as a whole significantly affect the dependent variable and vice versa.

3.7.3 Multiple Determination Coefficient Test (R^2)

The coefficient of determination (R^2) aims to find out how far the model capability describes the dependent variable being calculated. The value of R^2 is small / close to zero, meaning the ability of the independent variables in explaining the dependent variable is very limited or small. A large R^2 value approaches 1, meaning that the independent variables provide almost all the information needed to predict the variables of the bound variables.